AD-A061 160 SYSTEMS RESEARCH LABS INC DAYTON OHIO F/G 14/5
IMAGERY INTERPRETER PERFORMANCE IN THE COMPARISON OF SUBJECTIVE--ETC(U)
1977 E L GLIATTI, W L MARTIN, G G KUPERMAN F33615-75-C-0127

UNCLASSIFIED AMRL-TR-77-38 NL

1 OF 1
AD
A061160

END
DATE
FILMED
1-79
DDC

LEVEL

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

18 19

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AMRL-TR-77-38	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): IMAGERY INTERPRETER PERFORMANCE IN THE COMPARISON OF SUBJECTIVE ESTIMATES OF PHOTOGRAPHIC IMAGE QUALITY,	5. TYPE OF REPORT & PERIOD COVERED: Proceedings paper
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Edward L./Gliatti (AFAL); Wayne L./Martin (AMRL); Gilbert G./Kuperman (SRL, Inc.)	8. CONTRACT OR GRANT NUMBER(s): in part under F33615-75-C-0127
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Joint Study. Air Force Avionics Laboratory and Aerospace Medical Research Laboratory, AMD, AFSC, Wright-Patterson Air Force Base, Ohio 45433	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62202F, 7184-11-21, 2004-10-02
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Avionics Laboratory, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio 45433	12. REPORT DATE: 1977
	13. NUMBER OF PAGES: 19
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

6 2 10 15 16 11 17 12 21p.

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

DDC RECEIVED OCT 10 1978 F

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for release to the public; distribution unlimited

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Image quality
Photo interpreter performance
Image assessment
Tribar targets

340 400 LB

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

State of the art research in image quality assessment has been oriented toward objective measures of image quality, requiring microdensitometers and computers. Cost and time constraints in the operational situation, however, place emphasis on man dependent methods. This paper describes the performance of the image interpreter in a study comparing two of the more widely accepted Air Force subjective measures of image quality: tribar target resolution reading and visual edge matching. These techniques are described, interpreter certification is

#20

discussed, data derived from the application of each technique to a common imagery set are presented, and a comparison of the two methods reported.

B

IMAGERY INTERPRETER PERFORMANCE IN THE COMPARISON OF SUBJECTIVE ESTIMATES OF PHOTOGRAPHIC IMAGE QUALITY

Edward L. Gliatti
Dynamics and Environmental Evaluation Branch
Air Force Avionics Laboratory
Wright-Patterson AFB, Ohio 45433

Wayne L. Martin
Visual Display Systems Branch
Aerospace Medical Research Laboratory
Wright-Patterson AFB, Ohio 45433

Gilbert G. Kuperman
Special Projects Operation
Systems Research Laboratories, Inc.
Dayton, Ohio 45440

State-of-the-art research in image quality assessment has been oriented toward objective measures of image quality, requiring microdensitometers and computers. Cost and time constraints in the operational situation, however, place emphasis on man-dependent methods. This paper describes the performance of the image interpreter in a study comparing two of the more widely accepted Air Force subjective measures of image quality: tribar target resolution reading and visual edge matching. These techniques are described, interpreter certification is discussed, data derived from the application of each technique to a common imagery set are presented, and a comparison of the two methods reported.

INTRODUCTION

Background

Aerial photography represents a major source of information in both remote sensing and military reconnaissance/mapping applications.

To achieve the maximum information, the highest quality imagery needs to be produced for interpretation. The United States Air Force, in 1972, instituted an image quality control program, under the nickname SENTINEL SIGMA. Its purpose is to provide standardization and quality assurance capabilities to all USAF reconnaissance and mapping programs. The initiator of SENTINEL SIGMA (Crane, 1976) described the expected results as: "insuring that the maximum exploitation capability is available to intelligence analysts from all systems, while providing appropriate evaluation criteria to monitor and analyze the entire system from sensor performance through imagery exploitation."

In order to accomplish the SENTINEL SIGMA objectives, a Sensor Evaluation Center (SEC) was established at the Air Force Avionics Laboratory (AFAL), and is being assisted by the Aerospace Medical Research Laboratory (AMRL) in related human performance studies. Both of these activities are located at Wright-Patterson Air Force Base, Ohio. The mission of the SEC was set forth in Air Force Regulation (AFR 96-1), "Evaluation and Quality Assurance for U.S. Air Force Reconnaissance Imaging Systems."

An image evaluation workshop was held at AFAL in December 1975. Virtually every major Department of Defense reconnaissance or mapping organization was represented. Three major points were developed which summarized the state of the art in image quality assessment:

1. Tribar targets are relied on, by the operational commands, when they are available.

2. Concern exists that subjective measures of image quality do not produce agreement among users.
3. A need exists for proven techniques that are fast, simple and economical and that can be applied away from the laboratory environment.

Tribar Targets

Military Standard 150A, "Photographic Lenses," provides for the evaluation of lens/imaging characteristics against a standardized stimulus, the tribar target. The target is described as follows:

> The standard target element shall consist of two patterns (two sets of lines) at right angles to each other. Each pattern shall consist of three lines separated by spaces of equal width. Each line shall be five times as long as it is wide.

Successive patterns decrease in line (bar) width in a constant proportion, usually according to the sixth-root-of-two (1.12). A sufficient number of patterns, and range in bar widths, is provided to cover the requirements of the lens-film combination undergoing test.

Resolving Power

The dependent measure estimated through the exploitation of tribar targets is termed resolving power (RP). With respect to the tribar target, Military Standard 150A defines RP as the "ability to image closely spaced objects so that they are recognizable as individual objects" and their measurement as "the reciprocal of the center-to-center distance of the lines that are just distinguishable in the recorded image." The unit used to express RP data is cycles per millimeter (cy/mm) where one cycle corresponds to twice the bar width. Discussions of RP and its application as an image assessment technique are provided by Katz (1963), Brock et al. (1963, 1966), Pittman (1965), Charmin and Olin (1965), Attaya et al. (1966), Brock (1966, 1970), Mayo (1968), Noffsinger (1970), and Dainty and Shaw (1974). In general, these authors reported significant individual differences between readers, significant reader by target interactions, and the lack of a standard training methodology and criterion.

Visual Edge Matching

In recent years, a new image quality assessment technique, visual edge matching (VEM), has been proposed for application. It offers an obvious advantage over RP estimates in that no specially configured tribar target is required. Images of randomly occurring edges are compared against a reference matrix of calibrated edge images. Calibration is in cy/mm and the VEM technique is directly relatable to RP readings. In addition to the matrix, a more complex viewing station is required. A laboratory evaluation of the VEM technique has not been heretofore reported.

Purpose

The intent of this paper is to report on a comparison of RP and VEM performance estimates. The nature of this comparison is in the form of a correlation between two operator dependent image assessment methods. The VEM technique, while presumably requiring similar visual and cognitive processes, frees the evaluation from dependence on special targets.

This comparison is presented in the context of a well controlled aerial camera flight test evaluation. The unique training and professional experience represented by the subject set is of particular interest.

METHOD

Subjects

Thirteen males and one female participated in the RP portion of the experiment. All had normal or corrected 20/20 vision. Five of these same subjects served in the VEM portion of the experiment.

RP Training

Each reader successfully completed a training and certification program designed by the Defense Intelligence Agency (DIA). In this program the concepts of tribar target design and RP were presented, and the importance of the reader's work was explained. In addition, motivational training was provided to establish a criterion of "reasonable confidence" (i.e., less than absolute certainty). Sixty

paper prints containing imaged tribar targets for which RP values had been established were used to train the readers to criterion performance. The prints were read, in blocks of 15, each day by each reader until his mean reading error, for each pattern orientation, was not more than 6% and the standard deviation of the differences (between his reading and the established value) was less than 16%. In addition, three to four days were required to achieve criterion performance on a set of 34 glass plates for which RP values had also been established. Each plate contained a tribar target image to demonstrate the transfer of criterion performance from the training set to other tribar target imagery. Certification was based on the variability tolerances previously described.

VEM Training

A one-day familiarization program was provided. Subjects then worked with a set of training images (i.e., of identified VEM matrix element equivalents) until they demonstrated an absolute reading difference of no more than one step (nominally 12%). Criterion performance was then demonstrated using a set of test images with the "school solution" unknown to the subject. At present there is no certification procedure for VEM performance.

Subject Experience

Nine of the subjects were from a facility which employed RP and other (non-VEM) techniques as a major activity. Two subjects were from another facility in which VEM is the primary technique used and image quality estimation a major job function. The remaining three subjects resided in a laboratory in which both RP and VEM techniques are used and image quality assessment represented a significant

portion of their duties. All 14 subjects were certified RP readers. The five subjects used in the VEM experiment were from the two facilities using the VEM as a standard technique in their daily work. All readers can be considered experts and the best available, having enormous amounts of similar reading experience.

Aerial Photography

The aerial photography used in this experiment was acquired by a camera system undergoing sensor improvement flight testing at Edwards AFB, California. The camera combines the time interval between successive frames with a fore/aft nodding motion to obtain overlapping stereo coverage of the ground. The oblique pointing of the camera optical axis was through a small angle, and a near-vertical perspective photograph was obtained; the geometric distortion introduced was less than one percent. Four camera modes resulted from the combination of fore/aft pointing and target pattern orientation.

Targets

Fourteen Military Standard 150A tribar targets, located at Edwards AFB, were used for both RP and VEM evaluations. In each case, the target was imaged directly beneath (i.e., at nadir) the overflying aircraft.

The presence of multiple targets of identical spatial and spectral characteristics was intended to facilitate the collection [illegible]ry for RP collection by minimizing flight activity. Since these [illegible]gets were acquired on one pass, confounding due to changes in sun angle and atmospheric conditions was also minimized.

The camera obtained pairs of images of the ground scene, and because of overlapping coverage, 28 photographs were obtained. Since tribar patterns are oriented orthogonally to each other, a total of 56 observations were available. (The next to largest bar in each orientation was used for making the VEM comparions, thus preserving the same number of observations.)

Apparatus

All RP measurements were made at a light table equipped for variable illumination. A biocular microscope, providing variable magnification (to at least 90 diameters) was used for all readings.

All VEM measurements were made at a Visual Edge Match Comparator (Itek Corporation) with the same reference edge matrix. This instrument provides for matrix indexing in contrast and sharpness, separate light intensity controls for the matrix and light table channels, and a split field, double microscope comparator equipped for image rotation and separately adjustable magnification.

Procedure

Reading trials were self-paced. Subjects were permitted rest periods at their discretion. All targets were read in the order in which they occurred on the film.

RP Readings. The subject seated himself at the light table. The target image to be read was located. Light table illumination, magnification, and focus were adjusted with complete freedom by each subject for each reading in order that he have maximum self-confidence in his performance. All readers began with the largest

tribar pattern in each orientation and read in the direction of the smallest pattern. The RP reading criterion was as specified in the DIA standard on tribar target reading. To judge that a pattern had been resolved, the reader had to be reasonably confident that three bars had been present in the ground scene; that the bars were approximately equal in length; that there was a perceived contrast, along the entire length of each bar, between the bar and its surround; and that if a pattern was not judged to have been resolved, it was not followed by more than a single, smaller pattern which was judged to be resolved.

VEM Readings. The subject seated himself at the VEM Comparator. The target image to be read was located. The edge to be read (i.e., the next to largest bar of the tribar target) was located. Fine vertical and horizontal translation controls and the optical image rotator control were adjusted until the edge was aligned perpendicular to the split field dividing line. Magnification and focus were set for each microscope so that the density change across the edge was apparent and both microscopes were at equal magnification. The highest image sharpness row of the reference matrix was scanned, while adjusting both the matrix and light table brightness levels, until the best matching reference edge (in contrast) was determined and the brightness levels of each half of each edge were judged to match. The matrix was then searched in image sharpness, maintaining contrast and brightness, until the best visual matching reference edge was located.

Data Collection. The subject recorded the number of the smallest tribar pattern judged to have been resolved, or the number of the

reference matrix edge which best matched the subject image. The tribar targets were converted to cy/mm RP values based on the scale of the photograph and the bar width of the smallest pattern resolved. VEM matrix edges were calibrated against these tribar target images, recorded on the same film and processed under the same control parameters as the photography being read.

Experimental Design

The flight test program which acquired the imagery was designed to address the performance of the camera system. Additionally, the test range used in the program was unique in that multiple (14) tribar targets were available to facilitate the collection of imagery for resolving power estimation. Personnel directing the tests were interested in the performance of the camera, in each mode, and in the effect on camera performance caused by possible differences between the nominally identical target installations.

The major design was a two-factor, repeated measures analysis of variance (ANOVA) with modes and targets being the factors (Winer, 1962). To compare the two image quality assessment techniques with particular regard to the sensitivity of each technique to reader variability, a one-way ANOVA (Guilford, 1965) was used to assess subject differences.

RP Baseline

The objectives were to estimate camera performance by mode and to determine if the multiple target array truly represented identical stimuli. Bartlett's Test for Homogeneity of Variance (Snedocor and Cochran, 1967) was applied to the data from all 14 subjects. The

result indicated that a transformation was required, and following application of the tests for transformation described by Kirk (1968) a logarithmic transform was found to be most appropriate. A two-fac repeated measures ANOVA was applied to the transformed data. The summarized results of this test are shown in Table 1.

TABLE 1. TWO-FACTOR ANOVA SUMMARY: TRIBAR [14 Subjects; X' = Log (X + 1)]

Source	Subjects	DF	MS	F
Between Subjects	.9901	13		
Within Subject(s)	9.5661			
B (Modes)	1.8955	3	.6318	113.5*
BXS	.2171	39	.0056	
C (Targets)	2.1211	13	.1631	57.8*
CXS	.4767	169	.0028	
BC	4.0079	39	.1028	61.4*
BCXS	.8483	507	.0017	

*p < .01

As can be seen in Table 1, camera modes were significantly different from each other, targets were significantly different, and there was a significant interaction between modes and targets (all at $p < .01$).

RP Analyses

Since only five of the 14 subjects read both tribar and VEM imagery, a second two-factor ANOVA was performed using only the data from these subjects. The summary of this ANOVA is presented in Table 2. Again, camera modes, targets, and the mode by target interacti were all statistically significant ($p < .01$).

In analyzing the two sets of data (one with 14 readers, and the other with five) a Newman-Keuls multiple comparison test

(Kirk, 1968) was used with α set at .05. This procedure identified the specific mode and target elements which led to the significant findings in each of these ANOVAs. For modes, the result was identical in both analyses: no difference was found between Mode 1 and Mode 3. For targets, almost identical ordering and almost identical significant differences were found between the two data sets.

TABLE 2. TWO-FACTOR ANOVA SUMMARY: TRIBAR [Five Subjects; X' = Log (X + 1)]

Source	Subjects	DF	MS	F
Between Subjects	.6886	4		
Within Subject(s)	3.3440			
B (Modes)	.5777	3	.1926	16.19*
BXS	.1427	12	.0118	
C (Targets)	.6921	13	.0532	12.39*
CXS	.2235	52	.0043	
BC	1.3748	39	.0353	16.50*
BCXS	.3332	156	.0021	

*p < .01

Subject Differences--RP

All subjects who participated in this study had received training to the point of successfully demonstrating criterion performance in reading calibrated tribar targets. It was of interest, then, to investigate possible differences between subjects in this RP imagery set. This was done by means of a one-way ANOVA on modes. Table 3 presents the summary for this analysis. Again, the Newman-Keuls test (α = .05) was used to isolate the cause of the significant between-subjects effects. Subjects 2 and 3 were found to be different.

TABLE 3. ONE-WAY ANOVA SUMMARY: TRIBAR
[Five Subjects; X' = Log(X+1)]

Source	Subjects	DF	MS	F
Between Subjects	.0492	3	.0164	5.09*
Within Subjects	.0515	16	.0032	
Total	.1007	19		

*p < .05

VEM Analyses

The five subjects, as indicated above, read the second largest bars of the tribar targets using the VEM Comparator. Bartlett's Test was again applied to the raw data but no transformation was required. A two-factor, repeated measures ANOVA was applied to these data. Table 4 presents the ANOVA summary.

TABLE 4. ANOVA SUMMARY: VEM
(Five Subjects)

Source	Subjects	DF	MS	F
Between				
Subjects	59.1929	4		
Within				
Subject(s)	493.6071			
B(Modes)	45.5429	3	15.1810	11.34*
BXS	16.0643	12	1.3387	
C(Targets)	118.7000	13	9.1308	11.75*
CXS	40.4071	52	.7771	
BC	164.9571	39	4.2297	6.11*
BCXS	107.9357	156	.6919	

*p < .01

Camera modes were different from each other, targets were different from each other, and the mode by target interaction was significant (all at $p < .01$). Newman-Keuls comparisons were made between modes and between targets. Mode 2 was found to be different from the other three modes. Generally, the ordering of the targets was the same as found by the analysis of the RP data, but 17% fewer of the pairs (41 versus 48) were found to be significantly different from each other.

Subject Differences--VEM

A one-way ANOVA was applied to determine differences between the subjects. The result of this analysis is presented in Table 5.

TABLE 5. ONE-WAY ANOVA SUMMARY: VEM (Five Subjects)

Source	Subjects	DF	MS	F
Between Subjects	4.2281	3	1.4094	5.12*
Within Subjects	4.4005	16	.2750	
Total	8.6286	19		

*$p < .05$

The Newman-Keuls comparison ($\alpha = .05$) was made between pairs of subjects. Again, subjects 2 and 3 were found to differ.

Regression Analysis

One of the objectives of this study was to calibrate the VEM reference matrix against the RP readings obtained in the flight test. A regression analysis was performed using the two data sets generated

by the same five subjects. The RP readings were first normalized to a two-to-one target contrast, as suggested by Mayo (1968), then logarithmically transformed and pooled over subjects. The VEM scores were also pooled over subjects. The correlation coefficient obtained was -0.834 ($p < .01$). A correlation coefficient of -0.833 was obtained using the raw RP data, pooled over subjects. The coefficient is negative because reversed conventions were used to identify tribar patterns and VEM matrix edges.

The linear regression equation for the line of best fit is:

$$\mathrm{Log}(RP_{VEM}) = 2.04268 - 0.07331\,(VEM - 11.79286)$$

where

VEM = observed VEM matrix reading

RP_{VEM} = equivalent RP reading

CONCLUSIONS

Both the RP estimates and the VEM readings yielded essentially the same information concerning camera behavior. An additional mode difference was demonstrated from analysis of the RP data. The linear regression equation which was found to relate RP and VEM readings serves as a calibration of the VEM matrix for the film/processing combination used in the test. The relatively high correlation coefficient between the two techniques speaks well for the application of VEM as a stand-alone image quality estimator, and as an attractive candidate for satisfying the requirements of the 1975 image evaluation workshop.

The use of supposedly identical RP targets, intended to expedite the collection of images for RP reading, also leads to a confounding of system performance data. This is demonstrated by the significance of both targets and mode by targets interaction. To some extent, this can presumably be compensated for by applying the correction factor (Mayo, 1968), based on the measured aerial contrast ratio, to normalize target contrast across targets.

Despite the unusually high level of reader training and the demonstration of criterion performance, individual differences were found between interpreters. This finding presents questions regarding whether current training and certification procedures are a sufficient guarantee of reader uniformity. However, the

significant difference between readers occurred for representatives from two different facilities and, therefore, the question of the effect of specialized experience should also be addressed. The use of replicate readings is recommended as the basis for further study of these phenomena.

REFERENCES

Air Force Avionics Laboratory. Controlled Range Network (CORN) manual. Wright-Patterson AFB, Ohio: Air Force Avionics Laboratory, 15 September 1973.

Air Force Avionics Laboratory. Minutes of the 3rd SENTINEL SIGMA image evaluation workshop. Wright-Patterson AFB, Ohio: Air Force Avionics Laboratory, 9-11 December 1975.

Attaya, W. et al. Study of image evaluation techniques. Wright-Patterson AFB, Ohio: Air Force Avionics Laboratory, Technical Report No. AFAL-TR-66-343, November 1966.

Brock, G. C. Image evaluation for aerial photography. London: Focal Press, 1970.

Brock, G. C., Harvey, D. I., Kohler, R. J., and Myskowski, E. P., Photographic considerations for aerospace (2nd ed.). Lexington, Massachusetts: Itek Corporation, 1966.

Crane, A. C., Jr. Air Force image quality control program. Proceedings of the 42nd Annual Meeting of the American Society of Photogrammetry. Washington, D.C., February 22-28, 1976.

Dainty, J. C. and Shaw, R. Image science. New York: Academic Press, 1974.

Defense Intelligence Agency, Standardized assessment and expression of tribar resolution. Washington, D.C., Defense Intelligence Agency (undated).

Guilford, J. P. Fundamental statistics in psychology and education. New York: McGraw-Hill, 1965.

International Business Machine Company. Scientific subroutine package (5th ed.). GH20-0205-4, Version III. New York: White Plains, August 1970.

Itek Corporation. Itek visual edge match comparator, operation and maintenance manual, Itek 71-8068-1. Lexington, Massachusetts: Itek Corporation, 20 October 1971.

Katz, Amrom H. On the informal history of image evaluation techniques and practice. Rand Paper P-2789, The Rand Corporation, September 1973.

Kirk, Roger E. Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks/Cole, 1968.

Mayo, James William III. Photographic resolving power of aerial reconnaissance lenses as a function of target modulation. Technical Report 28. Tucson, Arizona: The University of Arizona, 9 August 1968.

Noffsinger, Edward B. Image evaluation: Criteria and applications. S.P.I.E. Journal, Vol. 9, November 1970.

Snedocor, George W. and Cochran, William G. Statistical methods (60th ed.). Ames, Iowa: Iowa State University Press, 1967.

U.S. Goverment Printing Office, MIL-STD-105A, military standard: photographic lenses (including changes through 28 January 1963). Washington, D.C., 12 May 1969.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.